

Übung 7: Nichtparametrische Regression

Aufgabe 1

Von einem linearen Spline-Polynom $s : [a, b] \rightarrow \mathbb{R}$ zu den Knoten 1, 2 und 3 sei bekannt, dass $s(0) = 1$, $s(1) = 1,3$, $s(2) = 5,5$, $s(4) = 6$ und $s(5) = 6$ gilt. Berechnen Sie die folgenden Werte:

- $s(0,5)$
- $s(3)$
- $\int_2^4 s(x) dx$

Aufgabe 2

Gegeben sei das quadratische Spline-Polynom $s : [a, b] \rightarrow \mathbb{R}$ mit

$$s(x) = 1 + 0,65x + x^2 + (x - 1)_+^2 + 0,6(x - 2)_+^2.$$

Berechnen Sie die folgenden Werte:

- $s(1,5)$
- $s'(1,5)$
- $s''(2,2)$

Aufgabe 3

Zeigen Sie, dass sich der der Nadaraya-Watson-Schätzer für die Regressionsfunktion f eines nichtparametrischen Regressionsmodells $y = f(x) + \varepsilon$ als Schätzer für den bedingten Erwartungswert $\mathbb{E}[Y|X = x]$ herleiten lässt.

Hinweis: Verwenden Sie $\mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} y \frac{f_{(X,Y)}(x,y)}{f_X(x)} dy$ und schätzen Sie die beiden dafür benötigten Dichten mittels der Kerndichteschätzer

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad \text{und} \quad \hat{f}_{(X,Y)}(x, y) = \frac{1}{nhh'} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) K\left(\frac{y - y_i}{h'}\right).$$

Aufgabe 4

Laden Sie den Datensatz 'mcycle' aus dem Paket 'MASS' in R. Für die Entwicklung von Sturzhelmen soll untersucht werden, wie die Beschleunigung des Kopfes von der Zeit nach einem Unfall abhängt. An diese Daten soll das nichtparametrische Regressionsmodell

$$y = f(x) + \varepsilon \tag{1}$$

angepasst werden.

- Erstellen Sie zunächst ein klassisches lineares Modell und plotten Sie dieses zusammen mit dem Datensatz.
- Erstellen Sie nun eine Polynomiale Regression vom Grad $d = 3$. Erstellen Sie einen Korrelationsplot der erklärenden Variablen. Was fällt Ihnen hierbei auf? Welche Anpassungsmöglichkeiten gibt es für das benutzte Regressionsmodell?
- Führen Sie eine nichtparametrische Regression mit verschiedenen Methoden durch: Nächste-Nachbarn-Schätzer, Nadaraya-Watson-Regressionsschätzer mit Gauß-Kern, Lokale Polynomiale Regression vom Grad $d = 3$ mit Gauß-Kern und Kubischen Spline. Bestimmen Sie die optimale Anzahl an Nachbarn k , Bandbreite h bzw. Anzahl der Knoten K mittels des approximierenden Kreuzvalidierungskriteriums (52). Plotten Sie die Regressionen und ermitteln Sie anhand eines Validierungsdatensatzes eine Schätzung für den erwarteten quadrierten Prognosefehler. Analysieren Sie auch die Plots und zeigen Sie auf, in welchen Bereichen die Verfahren vorteilhaft sind. Überprüfen Sie außerdem die Verwendung von quantilsverteilten und äquidistanten Knoten für den kubischen Spline.

Hinweis: Sie können die built-in Funktionen: `knnreg(x,...)`, `ksmooth(x,...)`, `locpoly(x,...)` & `bs(x,...)`, verwenden.